BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 9: e180398
doi: 10.3897/biss.9.180398

OPEN ACCESS

Conference Abstract

Tracking Natural Science Objects and Their Physical and Digital Derivatives in DINA

Jonas Grieb[‡], James Macklin[§], David Shorthouse[§], Christian Bölling[|], Volker Lohrmann[¶], Michaela Grein[¶], Etta Grotrian[¶], Satpal Bilkhu[§], Claus Weiland[‡]

‡ Senckenberg – Leibniz Institution for Biodiversity and Earth System Research, Frankfurt am Main, Germany
§ Agriculture and Agri-Food Canada, Ottawa, Canada
| Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany
¶ Übersee-Museum Bremen, Bremen, Germany

Corresponding author: Jonas Grieb (jonas.grieb@senckenberg.de)

Received: 25 Nov 2025 | Published: 01 Dec 2025

Citation: Grieb J, Macklin J, Shorthouse D, Bölling C, Lohrmann V, Grein M, Grotrian E, Bilkhu S, Weiland C (2025) Tracking Natural Science Objects and Their Physical and Digital Derivatives in DINA. Biodiversity Information Science and Standards 9: e180398. https://doi.org/10.3897/biss.9.180398

Abstract

Provenance plays an important role in natural history collections, but capturing this information accurately, proved challenging for legacy digital collection management systems. DINA*[1] is being developed to address these limitations through a process-oriented data model that more effectively captures the complexity and context of provenance information (Bölling et al. 2022). DINA is an open-source, robust sample- and specimen-based collection management system in production and developed by an unincorporated international consortium of technologists and practitioners of the natural sciences (Glöckler et al. 2020). Its governance and data models help to foster the adoption of FAIR principles (Findable, Accessible, Interopable, Reusable) and to integrate objects across science domains. DINA's innovative "samplistic" data model records metadata on stepwise, hierarchical processes that generate physical and digital derivatives from parent material samples, accommodating complex real-life sample trajectories, for example:

1. A fossil specimen is acquired by a museum, which contains the remains of several organisms in hardened resin. Later, it is sawed into smaller pieces;

This is an open access article distributed under the terms of the CC0 Public Domain Dedication.

some pieces are stored under new catalog numbers, a tiny piece is sent for a destructive C14 (radioactive carbon) analysis (Fig. 1 a).

2. A naturally deceased individual of a mammal species is collected under a material transfer agreement (MTA). Later, subsamples like teeth, bones, and tissues undergo different preparation and preservation processes and are finally stored in specialized collections, each with its own identifier (scheme); a subsample might even be retrieved for sequencing. Strong provenance tracking is required to ensure that regulatory constraints like the MTA are consistently passed down to any derivatives.

3. A soil core is collected from a sampling site and registered together with observational metadata according to the MIxS (Minimum Information about Any Sequence) soil extension standard. Several subsamples are extracted before the remaining core is preserved. The subsamples are subjected to DNA extraction and sequencing to analyze environmental DNA (eDNA). The resulting sequences are then processed computationally to identify the species present (Fig. 1 b).

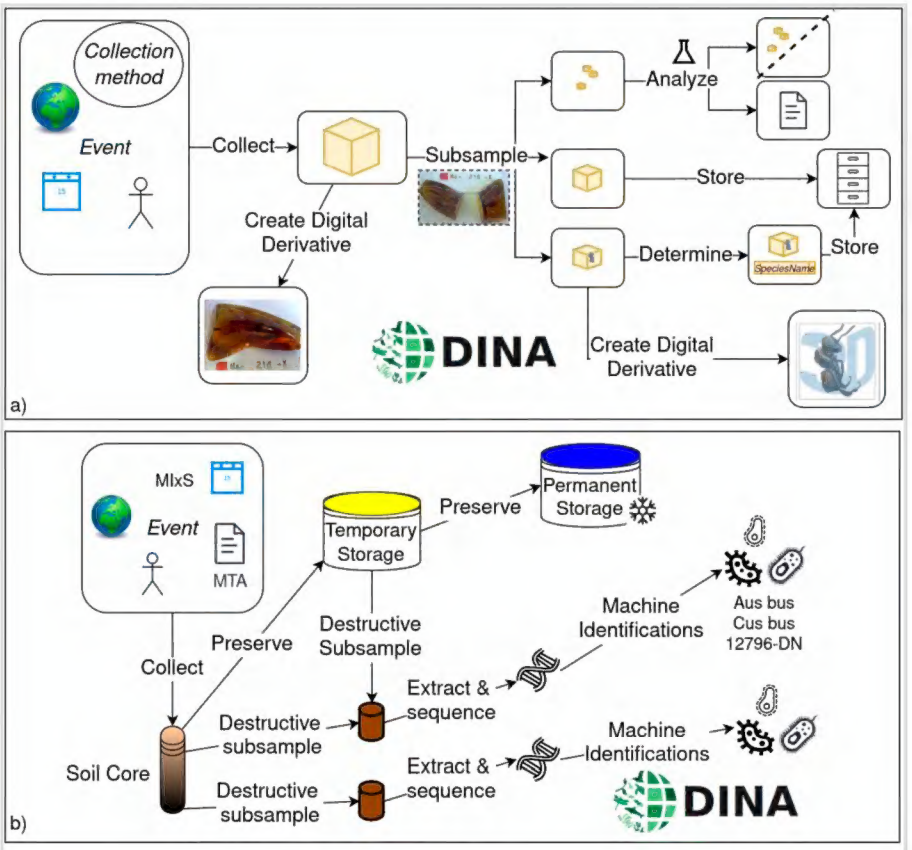

Figure 1.

DINA's provenance-preserving data model, visualized on an exemplary (a) fossil specimen lifecycle and (b) soil sample lifecycle. (Amber photos by M. Solórzano Kraemer under CC-BY-4.0, Icons by JGraph under CC-BY-4.0; http://icons8.com; https://iconsmind.com)

In all examples, the processing history of each subsample can be modeled in DINA so provenance can always be traced back to the original sample. Knowledge from any derivative analysis can be linked back to the parent sample and other derivatives. This ensures that regulatory restrictions applying to a parent material sample are passed down consistently. Besides parent-child relationships, the model can represent other

associations, e.g., host, parasite, and vector relationships linking samples in different collections. Each sample in the provenance chain, as well as first-class related objects like projects, collections, events, people, protocols, and storage, has a system-defined globally unique identifier. DINA integrates with external systems where possible via persistent identifiers, including reuse of scientific names from community-curated biodiversity sources through the Global Names Architecture.

DINA's application programming interface simplifies data import and migration and can be used by scientific programming languages like Python or R to access data objects or their relationships. Flexible user-defined "managed attributes" enhance object and derivative metadata when standards do not yet exist. When standards like MIxS or MIDS (Minimum Information about a Digital Specimen) exist, these are incorporated as field extensions.

Looking ahead, we aim to further strengthen linkages and provenance tracking in DINA, extending support to model complex bio-geo relationships and enhancing linkage of material samples to external resources such as publications.

Keywords

collection management system, material sample, provenance tracking, natural science collections, data model

Presenting author

Jonas Grieb

Presented at

Living Data 2025

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Bölling C, Bilkhu S, Gendreau C, Glöckler F, Macklin J, Shorthouse D (2022) Representation of Object Provenance for Research on Natural Science Objects: Samples, parts and derivatives in DINA-compliant collection data management. Biodiversity Information Science and Standards 6 https://doi.org/10.3897/biss.6.94531

- Glöckler F, Macklin J, Shorthouse D, Bölling C, Bilkhu S, Gendreau C (2020) DINA—Development of open source and open services for natural history collections & research. Biodiversity Information Science and Standards 4 https://doi.org/10.3897/biss.4.59070

Endnotes

*1 DIgital information system for NAtural history data - https://www.dina-project.net